Evaluation using Synthetic and Semi-Synthetic Biometric Data

Terrance E. Boult

El Pomar Professor of Innovation and Security
University of Colorado at Colorado Springs
and
CEO/CTO Securics

Presentation draws from a **decade** of work supported in part by DARPA HID, DHS SBIR, other DOD organizations
Currently Supported by ONR MURI and ONR STTR Tom McKenna

Securics®
Securics: The science of security with the protection of privacy™

UCCS™
University of Colorado at Colorado Springs

Outline

- Motivation and Definitions
- PhotoHeads
- 4D Photoheads
- SynFin Example and Issues
- Other Uses

Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Goals of Evaluation

- Show a particular system implementation meets requirements
- Show a particular system continues to operate as designed/tested
- Evaluate alternative system components
- Support research/design of new algorithm/sensor/system
- Understand the underlying "science"

Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Experimental Design

If your experiment needs statistics, then you ought to have done a better experiment.

Lord Ernest Rutherford (1871- 1937) English physicist. Nobel prize for chemistry 1908. **As quoted in N.Bailey. The Mathematical Approach to Biology and Medicine, Wiley, 1967.**

In Reality: Every experiment proves something. If it doesn't prove what you wanted it to prove, it proves something else. What is proves always depends on "statistics" whether you admit or not!

Experimental Design

Controlled Experiment (Hard science)

- Vary 1 or a few elements, hold all else constant.

Controlled Experiment (Social Science)

- Experiment where variable in question is varied between the test group and “control” group, with other variables balanced or randomized (e.g. RCT)

Natural Experiment

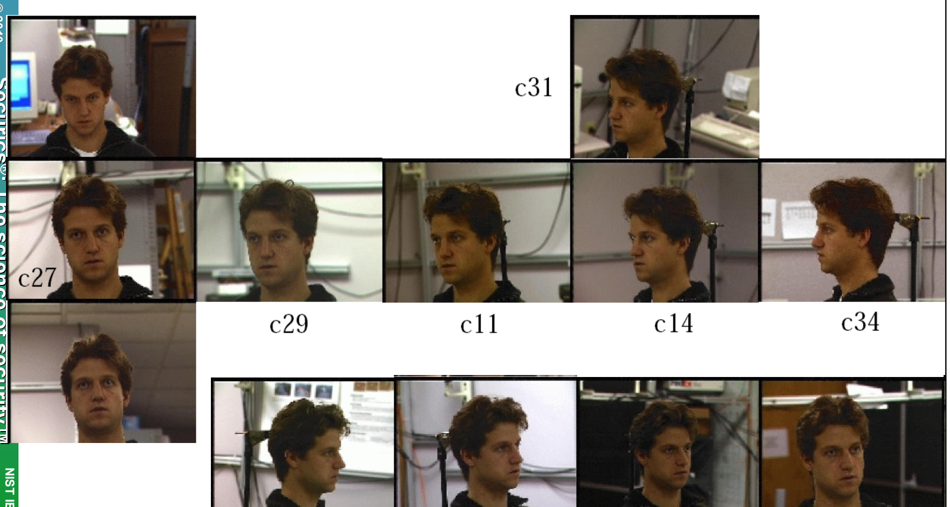
- Measurements from naturally occurring data, I.e. without formal controls group.

The greater the uncontrolled variation, the more data needed to reach a statistically relevant conclusion.

Control ←→ POWER

Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

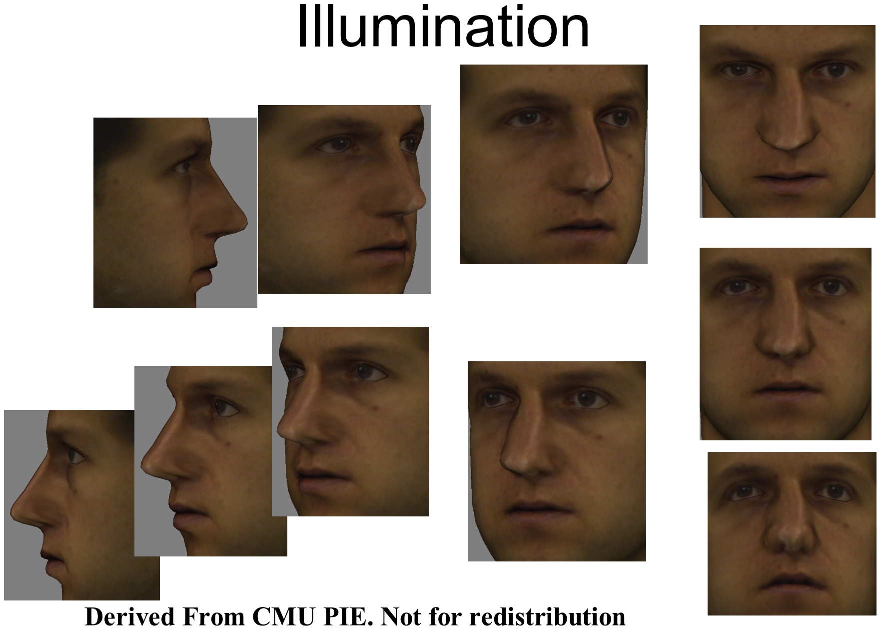
CMU PIE



The CMU Pose, Illumination, and Expression (PIE) Database
Terence Sim, Simon Baker, and Maan Bsat

Proc of the IEEE Int. Conf. on Automatic Face and Gesture Recognition, May, 2002.

PIE Controlled Pose, Lighting Illumination

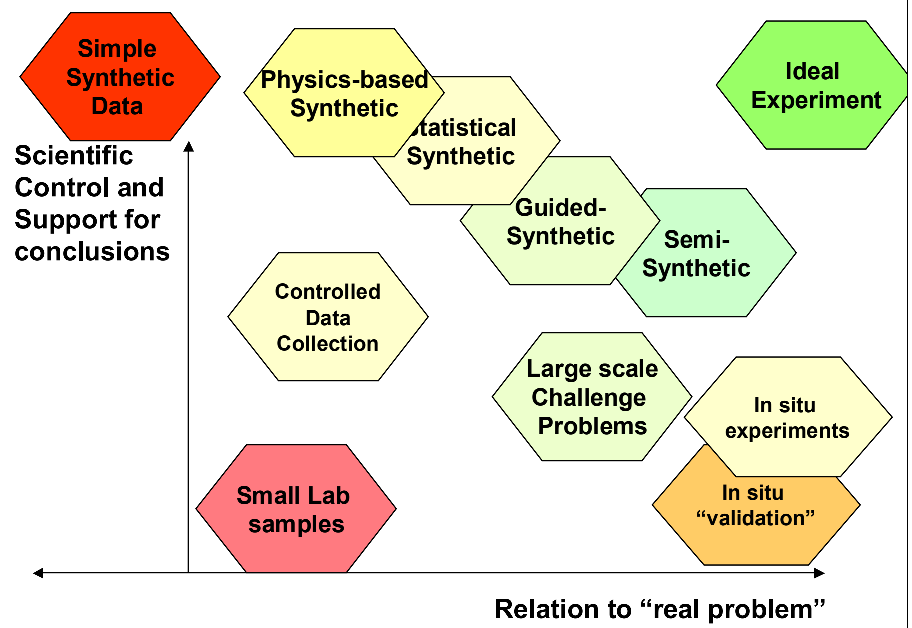


Derived From CMU PIE. Not for redistribution

Why (Semi-)Synthetic Evaluation

- Same say it is for more data, to build large datasets (at lower cost)
  - But this is limited by errors in “modeling”
- Maybe more important reason:
  - More experimental control!
  - Explore more conditions
- Less Obvious: testing assumptions

The “synthetic data” space



Definitions

- Synthetic (Pure Synthetic)
  - Driven by an (un-validated) generation model
- Modeled Synthetic
  - Driven by a generation model using parameters derived from and validated to real data.
- Guided Synthetic
  - Synthetic data where each sample is tied to real data.
- Semi-Synthetic
  - Real data mixed with artificial “sampling”
- Controlled Data Collection
  - Real data collected with controls on collection making it a Synthetic “Scenario/Operation”

Example “guided synthetic” vs real

Issue: how much do we consider regions outside ROI?

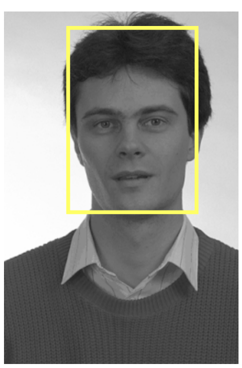
Our Goals

- Move to more and more and more automated/controlled/repeatable experiments.
- Build range of pure/guided/semi-synthetic
- Integrate with real system components (e.g. Real sensors, commercial algorithms, control/capture systems such as MBark)
- Domains long-range maritime biometric “evaluation”, multi-sensor multi-biometric fusion, adaptive fusion systems..

Synthetic "Evaluation" Validation

- Weak
  - Look at match/non-match distribution
  - Replicate known experiment on models as both probe and gallery?
- Increasing Levels of Validation/Testing:
  - Self-Image matching on ScreenShot
  - Replicate an known experiment on Screen-Captures using real gallery and synthetic probes.
  - Self-image matching based on Sensor Capture
  - Replicate an known experiment on Sensor-Captures using real gallery and synthetic probes.

Variations and Challenge



Cooperative Face

- Controlled pose
- Controlled position
- Controlled lighting

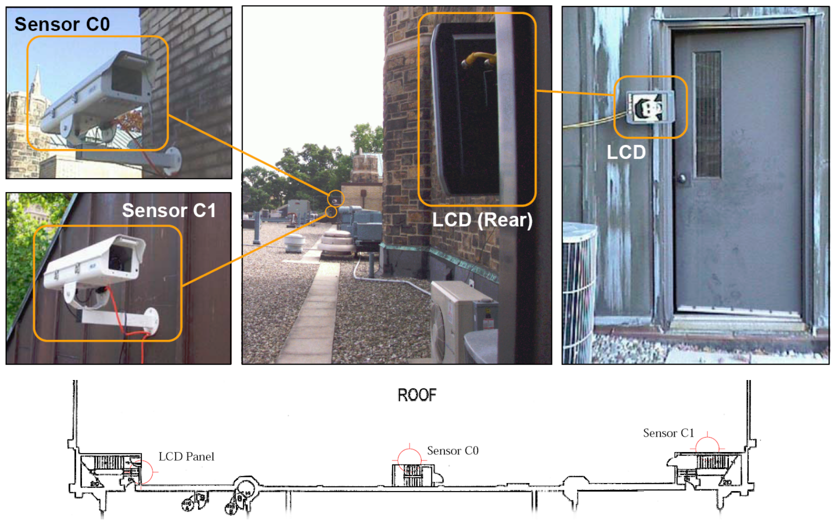
Non-Cooperative Face

- No control over subject
- Outdoors?
- Nighttime?

Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Early Photo-head Data Acquisition

Sensor : FOV 0.5° and 0.25° imaging (equivalent to 1600mm and 3200mm focal lengths).
Inter-pupil distance in resulting images is approx 120 pixels



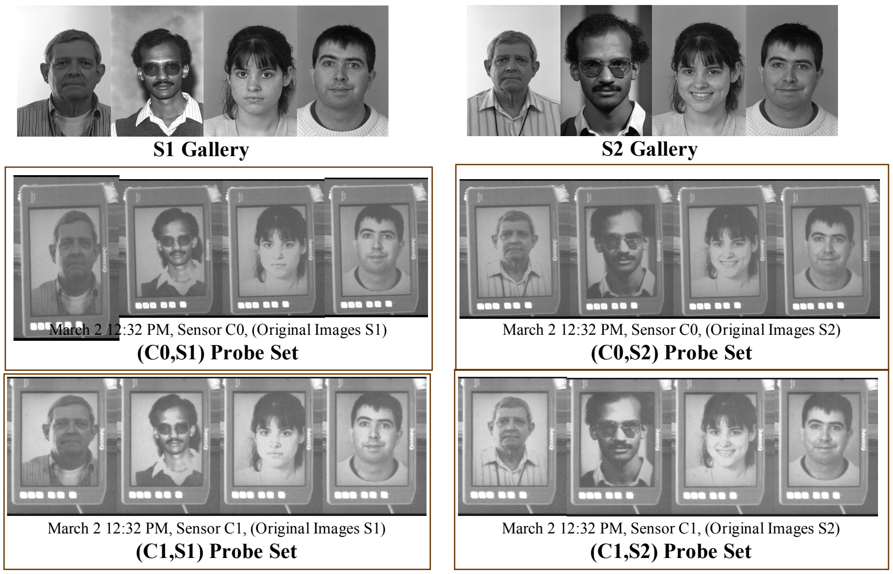
Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Photohead Elements

- Head/Face Models
- Imaging/Capture Systems
- Motion Models
- Lighting Models
- Display System (Not “real” system element)

Securics® : The science of security™ © 2010 NIST IBPC 2010 T. Boult UCCS

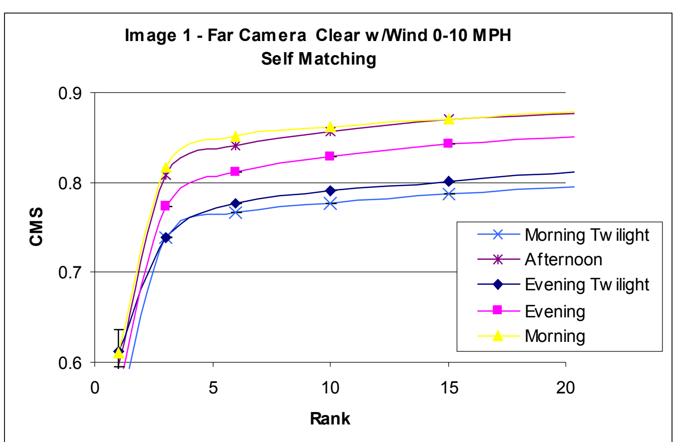
Example Early Photo-heads



Securics® : The science of security™ © 2010 NIST IBPC 2010 T. Boult UCCS

Examples Lessons with Early Photo-heads

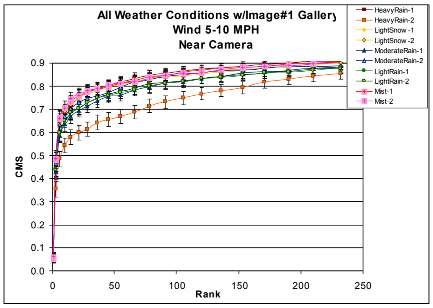
Variation over the day

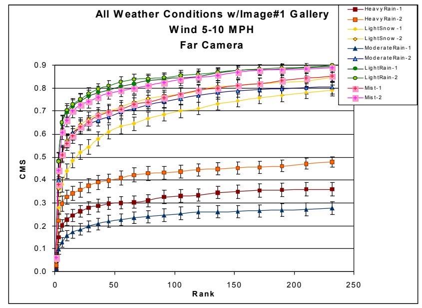


* leading commercial algorithm

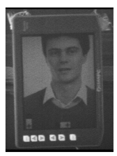
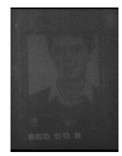
Impact of Weather

Weather





Rank 3 recognition about 50% (near) - much worse at 200ft

Clear  Snow  Rain 

Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

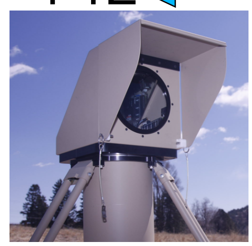
Early Photoheads results

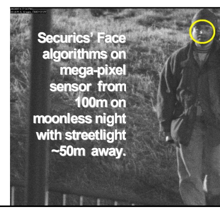
- Papers on Statistical Evaluation of System
- Lead to some non-obvious results
  - Multiple papers on Quality and System Failure Prediction
  - Performance enhancement via perturbations

Sensor System issues

- As if (unconstrained) face wasn't difficult enough…
- Choice of sensor has a huge impact on performance
  - Lighting – Depth of Field
  - Resolution – Motion Blur
  - Field of View

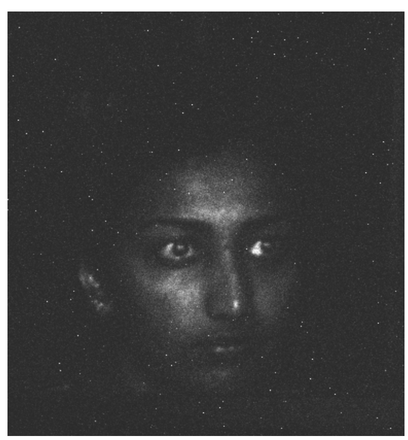
PTZ





Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Example "Dark Photo-heads"



Subset of CMU PIE, FERET data set re-imaged in a controlled, dark, indoor "photo-head" setting.
At Univ we have a 100m indoor "Dark room"

Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Qualitative Results

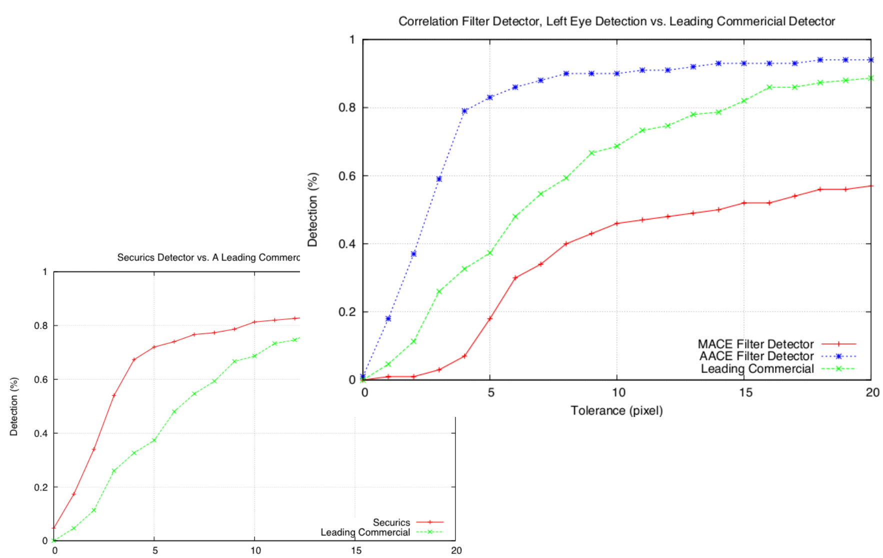
Securics Detector

A Leading Commercial Detector

Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Evaluating a lowlight detector

New detector vs. leading commercial detector



Motion Artifacts

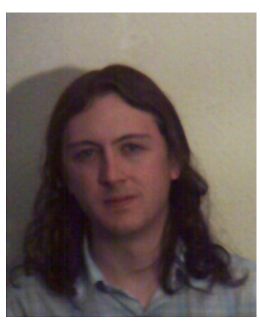
Typical motion blur

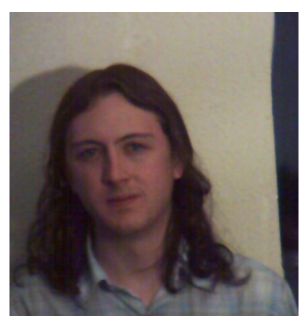
(~0.4 lux, yielding face lumens of 0.115 nits)

- Images taken approximately 100M from the EMCCD camera at dusk
- Top of the walking stride produces minimal blur

Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Other Artifacts





Obvious rolling shutter artifacts

- Affects Most/All CMOS sensors
- Even with a short integration time, the shutter is capturing data at different times for the top and bottom of the images

4D Photoheads
Animetrics Modeling

Securics® © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Securics® © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Securics®: The science of security™ © 2010 NIST IBPC 2010 T. Boult UCCS

Securics®: The science of security™ © 2010 NIST IBPC 2010 T. Boult UCCS

Ongoing "validation" results

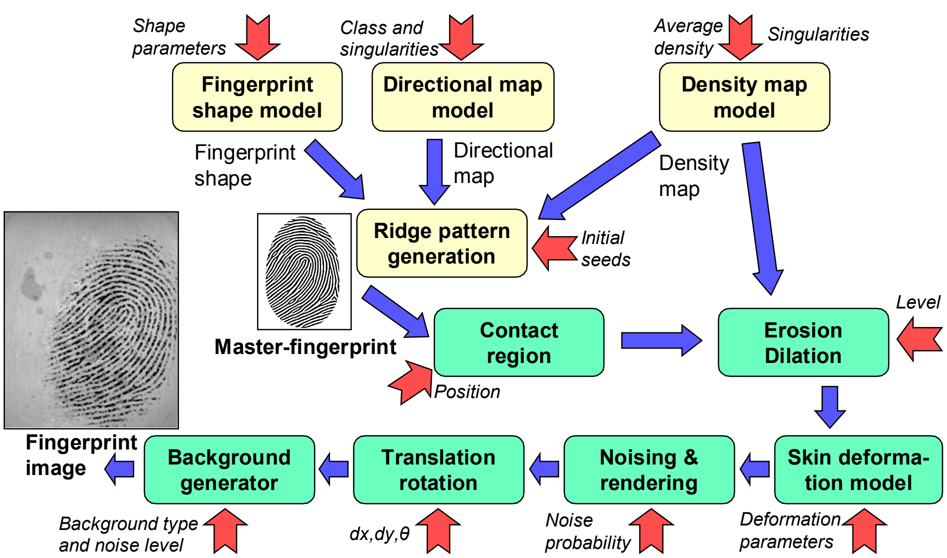
- FaceGen (Guided-synthetic):
  - Only 80-90% self-matching-screen used to build them! Only 46% on real test.
- Animetrics (Guided-synthetic):
  - 100% self-matching on multiple research and commercial recognition.
  - Replicated Frontal "PIE" results on real collected 4D Photoheads at 100m,200m
  - Working on full Guided-synthetic PIE
- Next Steps
  - Working on larger datasets.
  - Moving Platforms
  - Facial Surgery/Ageing/Weight Modeling

Securics® © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Outline

- Motivation and Definitions
- PhotoHeads
- 4D Photoheads
- SynFin Example and Issues
- Other Uses

How SFinGe works



In our view one of the best "synthetic" fingerprint systems. See R. Cappelli, D. Maio and D. Maltoni, "Synthetic Fingerprint-Database Generation", in Proc. *International Conference on Pattern Recognition (ICPR2002)*, Quebec City, August 2002.

How to validate the model?

- Can human's identify the synthetics?

About 90 people (many of them having a good background in fingerprint analysis) have been asked to find a synthetic fingerprint image among 4 images (3 of which were real fingerprints).
The synthetic image proved to be not distinguishable from the others

A B C D

Poll results	
A	23%
B	27%
C	21%
D	29%

Comparison of Curves

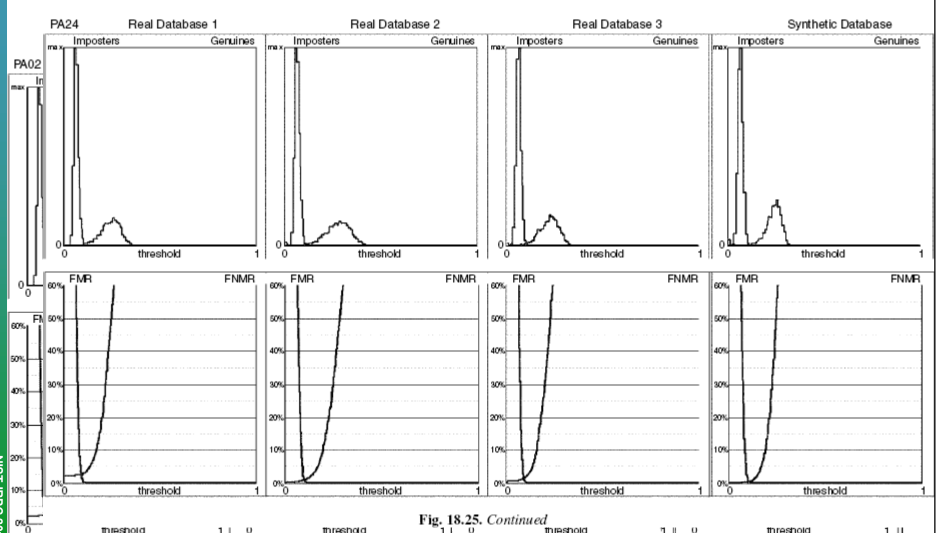


Fig. 18.25. Algorithms PA02 and PA24: impostor/genuine distribution and FMR/FNMR graph, for the four databases used in FVC2002.

They conclude it is "valid"

- They reasonably conclude it is about the same and real data and hence usable for testing. And for some testing it is!
- But biometric system performance and errors live in the per-match tails of distribution.
- Don't forget Weyman's talk this morning.. Everything in "experiment" matters

Actual Performance Differences

- FVC2004 real vs synthetic (DB4)
  - Looking at top 10 DB4 performers?
    Absolute Difference in Rank with DB1 is 7 positions!
  - Consider Relative Performance with Best Alg
    RERR_D = (AlgErr_D-BestErr_D)/BestErr_D)
  - Average Percentage Change in Relative ERR
    %CERR=(RERR_4-RERR_1)/RERR_4 = 323%
  - Average % change in Relative FMR100 = 156%
  - Average % change in Relative FMR1000=153%
  - For DB2 % R Changes were 74%, 186% 141%

Performance Differences vary

- FVC2004 is the "most" different as the "real" data had instructions to distort fingerprint. (It is one reason to normalize scores not use raw error rates)
- FVC2002 and FVC2006 are both closer (but still show differences between SFinGe and real data)
- FCV2002 DB1 vs DB4 has the following differences
  - Average change in absolute Range 2.7
  - Average Percentage Change in Relative ERR
    %CERR=(RERR_4-RERR_1)/RERR_4 = 159%
  - Average % change in Relative FMR100 = 57%
  - Average % change in Relative FMR1000=157%

Synthetic Fingerprint Issues

- Can we really conclude synthetic performance is the same?
- What is distribution of parameters
  - Type, minutia, ridge, orientation, pressure, moisture, system noise
- What biases are the models introducing?
  - Algorithms will tuned to these!

Other Uses of Semi-Synthetic

- Testing Assumptions (Micheals-Boult-08)
- Ongoing System "Revalidation"
- Validation of "algorithm" change
- Validation on component change
  - Sensors
  - Lenses
  - Bandwidth/Performance/compression…
- Hypothesized Variations (surgery, ageing, etc.)

Securics © 2010 Securics®: The science of security™ NIST IBPC 2010 T. Boult UCCS

Conclusions

- Defined different levels of "synthetic" data
- Experience has taught us LOTS of things can, and will, go wrong,
  go wrong,
  go wrong, …
  as you try to build (semi) Synthetic biometric evaluations.
- Semi-synthetic data offers experimental processes that can lead to new insights and, we believe, eventually better evaluations.